

AD-A080 899

CALIFORNIA UNIV BERKELEY OPERATIONS RESEARCH CENTER  
INFERENCE FOR THE EXPONENTIAL LIFE DISTRIBUTION. (U)

F/O 14/4

DEC 79 R E BARLOW, F PROSCHAN

APOSR-77-3179

UNCLASSIFIED

ORC-79-16

NL

ALL  
00000099

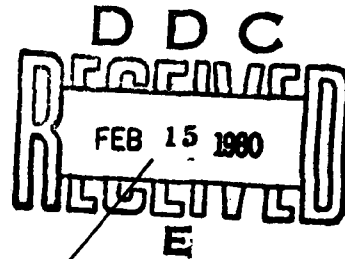
END  
DATE  
FILMED  
3 - 80  
100

ORC 79-16  
DECEMBER 1979

# INFERENCE FOR THE EXPONENTIAL LIFE DISTRIBUTION

by  
RICHARD E. BARLOW  
and  
FRANK PROSCHAN

**LEVEL**



ADA 080899

DOC FILE COPY

**OPERATIONS  
RESEARCH  
CENTER**

This document has been approved  
for public release and sale; its  
distribution is unlimited.

80 2 15 017

UNIVERSITY OF CALIFORNIA • BERKELEY

(12)

# INFERENCE FOR THE EXPONENTIAL LIFE DISTRIBUTION

by

Richard E. Barlow  
Department of Industrial Engineering  
and Operations Research  
University of California, Berkeley

and

Frank Proschan  
Department of Statistics  
Florida State University  
Tallahassee, Florida

This document has been approved  
for public release and sale; its  
distribution is unlimited.

DECEMBER 1979

ORC 79-16

This research was supported by the Air Force Office of Scientific Research (AFSC), USAF, under Grant AFOSR-77-3179 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER <b>14</b> ORC-79-16	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) <b>6</b> <u>INFERENCE FOR THE EXPONENTIAL LIFE DISTRIBUTION</u>		5. TYPE OF REPORT & PERIOD COVERED <b>9</b> Research Report
7. AUTHOR(s) <b>10</b> Richard E. Barlow <del>Frank</del> Proschan		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Operations Research Center University of California Berkeley, California <b>210 7-12</b>		8. CONTRACT OR GRANT NUMBER(s) <b>15</b> AFOSR-77-3179
11. CONTROLLING OFFICE NAME AND ADDRESS United States Air Force Air Force Office of Scientific Research Bolling AFB, D.C. 20332		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <b>16</b> 2394 <b>17</b> A51
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) <b>13</b> 46		12. REPORT DATE <b>11</b> Dec 1979
		13. NUMBER OF PAGES 44
		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Bayesian Life Testing Sampling Plans Failures and Total Time on Test		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  (SEE ABSTRACT)		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 0102-LF-014-6601

Unclassified **2705750** xlt  
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

## PREFACE

This report represents the first chapter of a book in preparation on inference and data analysis in reliability and life testing. The point of view adopted differs from that of most books on the subject in the following basic respect: *Prior information available to the reliability analyst is utilized fully in a formal statistical fashion.* Experience accumulated in helping engineers, quality assurance managers, scientists, biostatisticians, and others who must make estimates and reach decisions from either planned experiments or retrospective data has shown us that the point of view adopted throughout the book has resulted in useful solutions to real-life problems. By contrast, more classical statistical methods have often proven inadequate in many practical problems simply because the data available are insufficient to reach conclusions with a desired degree of assurance.

The book is intended primarily for actual use by the engineering and scientific practitioner, rather than for theoretical study and philosophical analysis by the statistician. Thus we omit a philosophical justification of the methods presented; rather, we rely on the fact that they have led to useful answers to problems that have arisen in practice.

One final point: Many of the methods and results are original and have not appeared in the literature. This fact has led us to issue the chapters as reports under our research grants.

- i -

Accession For	
NTIS GMA&I	
DDC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Availand/or special
A	

#### ABSTRACT

↓  
This paper constitutes Chapter 1 of a book on "Inference and Data Analysis for Reliability and Life Testing." The purpose is to provide an elementary coherent Bayesian foundation for research in statistical reliability theory. The results in Section 3 and the Appendix concerning the role of failures and total time on test in statistical inference for the exponential model is new.  
↑

# TABLE OF CONTENTS

	Page
1. BASIC CONCEPTS . . . . .	2
The Likelihood Function . . . . .	4
Bayes' Theorem . . . . .	5
Prior and Posterior Distributions . . . . .	6
Sufficient Statistics . . . . .	11
The Sample Space . . . . .	15
2. SELECTED LIFE TEST SAMPLING PLANS . . . . .	18
Sampling Plan (a). Complete Observation Until a Specified Number of Failures Have Occurred . . . . .	18
Sampling Plan (b). Observation Terminated at Fixed Time (Truncated Sampling) . . . . .	21
Sampling Plan (c). Inverse Binomial Sampling . . . . .	22
3. INFERENCE BASED ON MEAN LIFE . . . . .	26
The Mean of the Posterior Density . . . . .	27
Arbitrary Prior Density . . . . .	30
4. NOTES AND REFERENCES . . . . .	34
APPENDIX A: POSTERIOR DISTRIBUTIONS CORRESPONDING TO ARBITRARY PRIORS AND LIKELIHOODS WITH THE MONOTONE RATIO PROPERTY . . . . .	38

## INFERENCE FOR THE EXPONENTIAL LIFE DISTRIBUTION

by

Richard E. Barlow and Frank Proschan

Our objective is to develop methodology for analyzing life test data. Initially, we have only data--no mathematical models. Through an exploratory data analysis (see Chapter 3) or an analysis based on the physical processes generating the data, we may select an exponential life distribution model as appropriate for the analysis of the data. Of course, this may be at best only an approximation to the "true" underlying life distribution. With this reservation in mind, we suppose the life distribution  $F$  is exponential; specifically:

$$(0.1) \quad F(x \mid \lambda) = 1 - e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0,$$

where  $\lambda$  is the unknown constant failure rate. The vertical bar in  $F(x \mid \lambda)$  indicates that we are conditioning on the parameter  $\lambda$ ; i.e., for specified  $\lambda$  the distribution is exponential with failure rate  $\lambda$ . The corresponding density is:

$$(0.2) \quad f(x \mid \lambda) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$



# 1. BASIC CONCEPTS

To begin with, we suppose the life test data consist of observed complete lifetimes  $x_1, x_2, \dots, x_n$  on  $n$  units. For example, Table 1.1 lists ordered lifelengths of 100 Kevlar 49/Epoxy strands subjected to a high static load (Barlow, Toland and Freeman (1979)). An exploratory data analysis indicates that an exponential life distribution may be appropriate for analyzing these lifelengths. Thus, we assume that the observations constitute a sample of  $n$  independent, identically distributed random variables with distribution  $F$  given by (0.1). Although we assume a fixed  $\lambda$  exists which specifies  $F$ , we are uncertain as to the true value of  $\lambda$ , and seek a method which uses the data to express probabilistically our uncertainty regarding the true value of  $\lambda$ .

The first step is to evaluate the joint probability density of the random lifetimes  $X_1, \dots, X_n$  evaluated at the observed values  $x_1, \dots, x_n$ . Since we are assuming that  $X_1, \dots, X_n$  are independent given  $\lambda$ , the joint density of the observed values is:

$$(1.1) \quad \prod_{i=1}^n f(x_i | \lambda) = \lambda^n \exp \left( -\lambda \sum_{i=1}^n x_i \right).$$

TABLE 1.1  
TIMES TO FAILURE OF STRANDS SUBJECTED  
TO STRESS AT 80% OF MEAN RUPTURE STRENGTH

Rank	Hours	Rank	Hours	Rank	Hours	Rank	Hours
1	1.8	26	84.2	51	152.2	76	285.9
2	3.1	27	87.1	52	152.8	77	292.6
3	4.2	28	87.3	53	157.7	78	295.1
4	6.0	29	93.2	54	160.0	79	301.1
5	7.5	30	103.4	55	163.6	80	304.3
6	8.2	31	104.6	56	166.9	81	316.8
7	8.5	32	105.5	57	170.5	82	329.8
8	10.3	33	108.8	58	174.9	83	334.1
9	10.6	34	112.6	59	177.7	84	346.2
10	24.2	35	116.8	60	179.2	85	351.2
11	29.6	36	118.0	61	183.6	86	353.3
12	31.7	37	122.3	62	183.8	87	369.3
13	41.9	38	123.5	63	194.3	88	372.3
14	44.1	39	124.4	64	195.1	89	381.3
15	49.5	40	125.4	65	195.3	90	393.5
16	50.1	41	129.5	66	202.6	91	451.3
17	59.7	42	130.4	67	220.2	92	461.5
18	61.7	43	131.6	68	221.3	93	574.2
19	64.4	44	132.8	69	227.2	94	653.3
20	69.7	45	133.8	70	251.0	95	663.0
21	70.0	46	137.0	71	266.5	96	669.8
22	77.8	47	140.2	72	267.9	97	739.7
23	80.5	48	140.9	73	269.2	98	759.6
24	82.3	49	148.5	74	270.4	99	894.7
25	83.5	50	149.2	75	272.5	100	974.9

### The Likelihood Function

To focus attention on the parameter of interest  $\lambda$ , we regard (1.1) as a function of  $\lambda$  and call

$$(1.2) \quad L(\lambda \mid x_1, \dots, x_n) = \lambda^n \exp \left( -\lambda \sum_{i=1}^n x_i \right)$$

the *likelihood function*. (The likelihood, although a function of the parameter  $\lambda$ , is *not* a probability density in the parameter. Hence the vertical bar in  $L$  does *not* signify a conditional probability, as it usually does.) The likelihood function provides a means of quantifying the information contained in the data concerning the unknown true value of the exponential parameter  $\lambda$ .

Suppose a unique value  $\hat{\lambda}$  of  $\lambda$  exists maximizing the likelihood function; then we call  $\hat{\lambda}$  the *mode* of  $L(\lambda \mid x_1, \dots, x_n)$  and the *maximum likelihood estimator* (MLE) of  $\lambda$ . In general, the MLE, when it exists, is a very useful concept.

To simplify the calculation of  $\hat{\lambda}$ , we use the fact that the maximum of the likelihood, when it exists, is achieved at the same value of  $\lambda$  as is the maximum of the logarithm of the likelihood. Thus we compute

$$\frac{d}{d\lambda} \ln L(\lambda \mid x_1, \dots, x_n) = \frac{n}{\lambda} - \sum_{i=1}^n x_i,$$

and set the derivative equal to 0. We readily obtain

$$\hat{\lambda} = n \left( \sum_{i=1}^n x_i \right)^{-1},$$

and verify that  $\hat{\lambda}$  maximizes  $L(\lambda \mid x_1, \dots, x_n)$  for fixed  $x_1, \dots, x_n$ .

The MLE  $\hat{\lambda}$  may be a very satisfactory estimator of the unknown failure rate  $\lambda$  for moderate to large sample sizes  $n$ . (Caution: For more complex life distribution models involving an infinite number of unknown parameters, the MLE of the parameters may be quite misleading. See Marshall and Proschan (in process) for an example in which the MLE converges to the *wrong* set of the parameter values even though the sample size tends to infinity. Also see Lindley (1972), p. 12, for a similar two parameter example and Basu (1975), p. 34, for a one parameter example.) In the present case of estimation of the single parameter  $\lambda$  of the exponential, our uncertainty as to the true value of  $\lambda$  stems from the fact that our sample size  $n$  is finite.

A good way to express our uncertainty concerning  $\lambda$  is by means of a *probability distribution* for  $\lambda$ . To display explicitly this point of view, we let  $\tilde{\lambda}$  denote a random variable expressing our uncertainty concerning the unknown true value of  $\lambda$ .

### Bayes' Theorem

A key theorem based on this point of view is the fundamental Bayes' Theorem. It provides a method for computing the probability density of the random variable expressing our uncertainty concerning the parameter conditioned on the observed data.

#### 1.2 Theorem (Bayes' Theorem)

Let (a)  $X$  and  $\tilde{\theta}$  be random variables with joint probability density  $p(x, \theta)$ , (b)  $p(x | \theta)$  and  $p(\theta | x)$  denote the corresponding conditional densities, and (c)  $\pi(\theta)$  denotes the marginal density of  $\tilde{\theta}$ . Let  $(H)$  be the parameter space, i.e.,  $\theta \in (H)$ . Then

$$(1.3) \quad p(\theta | x) = \frac{p(x | \theta) \pi(\theta)}{\int_{\mathcal{H}} p(x | \theta) \pi(\theta) d\theta}.$$

Proof:

The joint density  $p(x, \theta)$  of  $X$  and  $\theta$  may be written as:

$$p(x, \theta) = p(x | \theta) \pi(\theta).$$

By definition of a conditional probability density,

$$p(\theta | x) = p(x, \theta) / p(x),$$

when  $p(x) > 0$ , where

$$p(x) \stackrel{\text{def}}{=} \int_{\mathcal{H}} p(x | \theta) \pi(\theta) d\theta.$$

By combining the three equalities in the steps just above, we immediately obtain the desired conclusion (1.3). ||

### Prior and Posterior Distributions

Before analyzing statistical data, it is helpful and efficient to assess relevant information. A convenient way to accomplish this is to formulate a probability density on the parameter(s) of the model selected. We realize that generally we cannot be certain that the model selected is the "true" model and that the prior distribution for the parameter(s) specified for the model is perfectly correct. However, once we select an appropriate model, and a prior distribution on the parameter space for that model, we may complete a useful and informative data analysis in an unambiguous fashion using only the standard calculus of probability theory.

### The Prior Density

First, we confine our choice of prior densities to *proper* densities; i.e.,  $\pi(\lambda)$  satisfies

$$\int_0^{\infty} \pi(\lambda) d\lambda = 1 .$$

Next, to motivate the concept of a natural conjugate prior for  $\tilde{\lambda}$ , we suppose that in the particular problem under discussion, we have very little prior information concerning  $\tilde{\lambda}$ . It seems natural to assume initially a rectangular prior density:

$$\pi_0(\lambda) = \begin{cases} M^{-1} & \text{for } 0 \leq \lambda \leq M \\ 0 & \text{otherwise,} \end{cases}$$

where  $M$  is a very large number (say  $M = 10^{12}$ ). Under this assumption on the prior density, we assign the same probability that  $\tilde{\lambda}$  is in any interval in  $[0, M]$  of a specified length. For example, the a priori probability that  $5 \leq \tilde{\lambda} \leq 10$  is the same as the a priori probability that  $19 \leq \tilde{\lambda} \leq 24$ .

### The Posterior Density

Suppose we have observed a sample of  $n$  lifelengths  $x_1, \dots, x_n$  having joint density  $p(x_1, \dots, x_n | \lambda)$ . By Bayes' Theorem, the *posterior density* of  $\tilde{\lambda}$  is given by:

$$\pi_1(\lambda | x_1, \dots, x_n) = p(x_1, \dots, x_n | \lambda) \pi_0(\lambda) / \int_0^{\infty} p(x_1, \dots, x_n | \lambda) \pi_0(\lambda) d\lambda .$$

Recall that the *likelihood*  $L(\lambda \mid x_1, \dots, x_n) = p(x_1, \dots, x_n \mid \lambda)$ , namely the probability density of the observed outcome considered as a function of the parameter  $\lambda$ . Thus we may write:

$$\pi_1(\lambda \mid x_1, \dots, x_n) = L(\lambda \mid x_1, \dots, x_n) \pi_0(\lambda) / \int_0^{\infty} L(\lambda \mid x_1, \dots, x_n) \pi_0(\lambda) d\lambda.$$

Since  $\lambda$  has been integrated out in the denominator, the denominator is now a constant with respect to  $\lambda$ . Hence

$$\pi_1(\lambda \mid x_1, \dots, x_n) \propto L(\lambda \mid x_1, \dots, x_n) \pi_0(\lambda),$$

where  $\propto$  means "proportional to." Thus the right hand side is the same as the left hand side up to a constant which does not depend on the parameter  $\lambda$ . Notice that since the data  $x_1, \dots, x_n$  have already been observed, the data are *not* considered variables at this stage of the analysis.

Assuming the rectangular prior  $\pi_0(\lambda) = M^{-1}$  for  $0 \leq \lambda \leq M$ , we obtain for the posterior density of  $\lambda$ :

$$\pi_1(\lambda \mid x_1, \dots, x_n) = M^{-1} \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) / M^{-1} \int_0^M \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) d\lambda.$$

For  $n^{-1} \sum_{i=1}^n x_i \gg M^{-1}$ , this is approximately:

$$(1.4) \quad \pi_1(\lambda \mid x_1, \dots, x_n) \doteq \left(\sum_{i=1}^n x_i\right)^{n+1} \lambda^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) / \Gamma(n+1),$$

where  $\Gamma(n+1) = \int_0^\infty u^n e^{-u} du$  is the gamma function. In computing (1.4) we have used the fact that  $\int_0^\infty c^{n+1} u^n e^{-cu} du = \Gamma(n+1)$  for all  $c > 0$ .

Thus, if we assume initially a rectangular prior on  $[0, M]$ ,  $M$  large, the resulting posterior density is approximately of the form:

$$(1.5) \quad b^a \lambda^{a-1} e^{-b\lambda} / \Gamma(a) \quad \text{for } \lambda > 0,$$

where  $a, b > 0$ . Note that the posterior density is approximately a gamma density with shape parameter  $a = n+1$  and scale parameter

$$b = \sum_{i=1}^n x_i.$$

Now, suppose we obtain an additional independent random sample of lifelengths  $y_1, y_2, \dots, y_m$ . Then it is reasonable to use as our *new* prior density, the posterior density (1.4) obtained from the previous sample. Using as our new prior:

$$\pi_1(\lambda) = b^a \lambda^{a-1} e^{-b\lambda} / \Gamma(a)$$

with  $a = n+1$  and  $b = \sum_{i=1}^n x_i$ , we obtain:

$$\pi_2(\lambda \mid y_1, \dots, y_m) = \frac{\lambda^m \exp\left(-\lambda \sum_{i=1}^m y_i\right) b^a \lambda^{a-1} e^{-b\lambda} / \Gamma(a)}{\int_0^\infty \lambda^m \exp\left(-\lambda \sum_{i=1}^m y_i\right) [b^a \lambda^{a-1} e^{-b\lambda} / \Gamma(a)] d\lambda},$$

or

$$(1.6) \quad \begin{aligned} & \pi_2(\lambda \mid y_1, \dots, y_m) \\ &= \left(b + \sum_{i=1}^m y_i\right)^{m+a} \lambda^{m+a-1} \exp\left(-\lambda \left(b + \sum_{i=1}^m y_i\right)\right) / \Gamma(m+a). \end{aligned}$$



Thus,  $a$  is increased by the additional number  $m$  of observed failures and  $b$  increased by the additional total time on test  $\sum_{i=1}^m y_i$  to obtain the new posterior density  $\pi_2$ ; note however, that the *form* of the posterior density, the gamma, is retained.

Because of this preservation property (the gamma prior used in the exponential model leads to a gamma posterior), the gamma prior is called the "natural conjugate" prior for the parameter  $\lambda$  in the exponential model. More generally, a conjugate prior distribution is "conjugate" with respect to a given statistical model if the form of the posterior is the same as that of the prior; the parameters of the posterior distribution will of course, change in accordance with the data observed. In the present case, we can interpret the prior density parameter  $a - 1$  (if  $a$  is an integer  $> 1$ ) as the number of observations in a previous experiment (actual or conceptual) and  $b$  as the corresponding total time on test.

In the present exponential model, the gamma prior for  $\lambda$  is mathematically convenient and has intuitive interpretation. However, the analyst is not confined to a choice with this family. Rather, the choice of the prior distribution should always reflect the best possible specification of the analyst's prior information concerning the unknown parameter. Thus, in reporting the results of the data analysis, the analyst should present his specification of the prior and the basis for his choice. In Chapter 3, we will discuss in more detail the assignment of the prior distribution.

### 1.3 Example

Table 1.1 lists the observed lifetimes of 100 organic fiber strands subjected to a static load of 8009 grams which corresponds to 80% of their mean rupture strength. Experience has shown that the lifetime

of an organic fiber strand at relatively high stress can be reasonably well fitted by an exponential life distribution. Thus, we assume

$$P[\text{Lifelongth} > x] = e^{-\lambda x} \quad \text{for } x \geq 0 ,$$

where  $\lambda$  is unknown. As described above, we calculated the MLE of  $\lambda$  as:

$$\bar{\lambda} = 4.78 \times 10^{-3} / \text{hour} .$$

Since the sample size of 100 is moderately large, the likelihood

$$L(\lambda \mid x_1, \dots, x_{100}) = \lambda^{100} \exp \left( -\lambda \sum_{i=1}^{100} x_i \right)$$

will override in importance the rectangular prior:

$$\pi_0(\lambda) = M^{-1} , \quad 0 < \lambda \leq M ,$$

when  $M \gg \bar{\lambda}$ . From the prior  $\pi_0$  we calculate the posterior density of  $\lambda$  to be approximately a gamma with parameters  $a = 101$  and  $b = 20,917$  hours. (Note that  $a$  is dimensionless while  $b$  is measured in hours). See Figure 1.1. The mode of the posterior density may be used to estimate  $\lambda$ . In Chapter 2 we discuss other estimators.

### Sufficient Statistics

By a *statistic*, we mean a function (possibly vector-valued) of the data. A statistic, of course, is often used to estimate an unknown parameter of interest. Clearly, the complete set of data observed is trivially a statistic. For a large set of data, working with *all* of the observations may be tedious or even unmanageable. Thus, we are motivated to find a statistic of smaller dimensionality like the sample mean (of

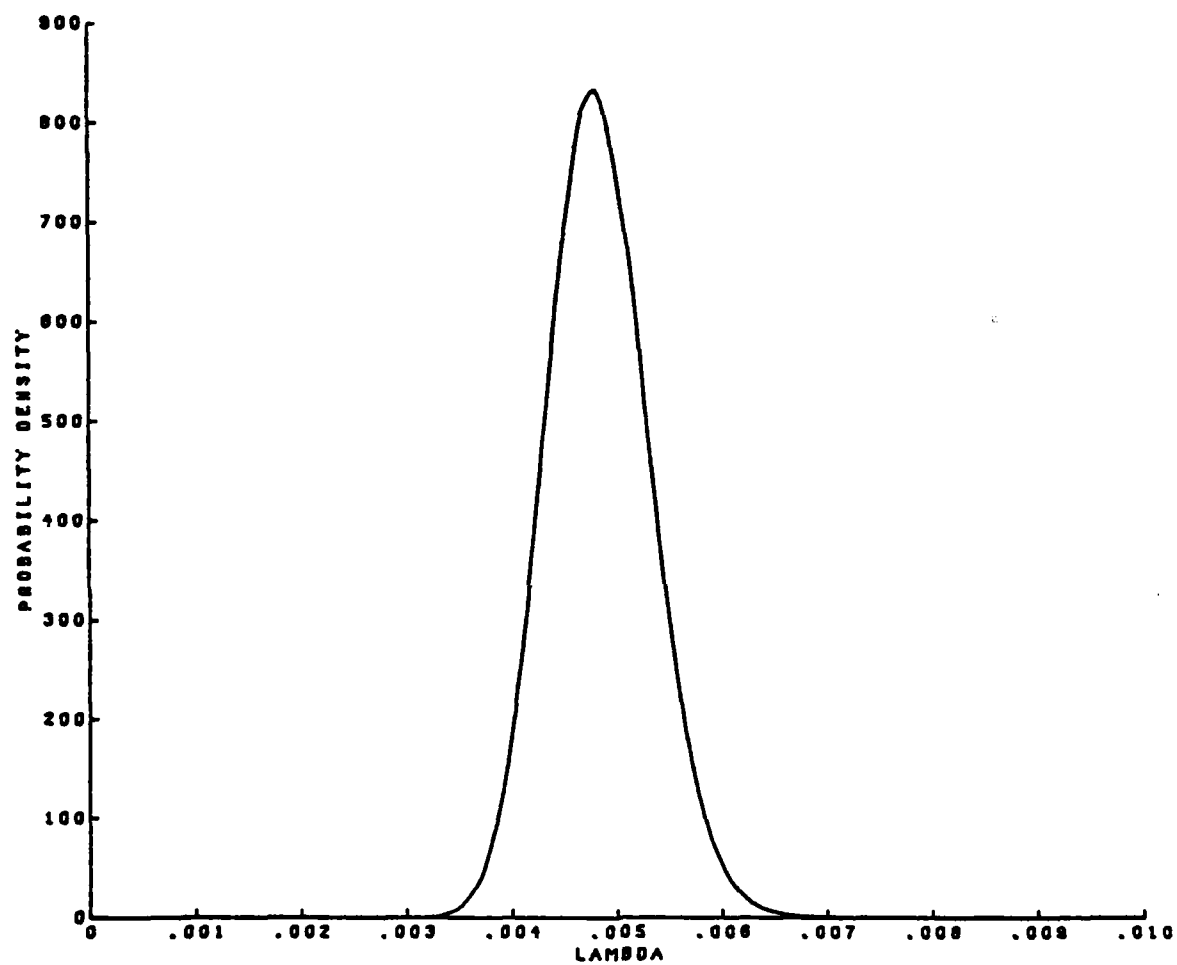


FIG 1.1  
COMPUTER PLOT OF POSTERIOR GAMMA DENSITY OF LAMBDA

dimension one) or the number of failures *and* the total time on test (of dimension two), but which contains all of the information in the sample concerning the parameter. Preferably, we would like to estimate the parameter using a statistic of lower dimensionality which summarizes all of the information in the sample. We will often use  $D$  to denote the observed data. For example,  $D$  could stand for the vector of observed values  $(x_1, x_2, \dots, x_n)$ . Later it will denote more complicated data sets. This motivates:

#### 1.4 Definition

Let  $D$  denote the data with probability density  $p(D | \theta)$  indexed by the parameter  $\theta$ . A statistic  $s$  is *sufficient* for  $\theta$  if and only if for every prior  $\pi(\theta)$ , the posterior

$$\pi(\theta | D) = p(D | \theta)\pi(\theta) / \int_H p(D | \theta)\pi(\theta)d\theta$$

depends on the data  $D$  only through the statistic  $s$ ; i.e., for every prior  $\pi$ , the posterior can be written as  $\pi(\theta | s)$ .

Intuitively, knowing  $s$  we are as informed about the parameter  $\theta$  as when we know all the data collected.

There may be several sufficient statistics available for estimating a parameter. Clearly, from among these we would prefer to make use of one of lowest dimensionality. For a given parameter, there actually may be more than one sufficient statistic of lowest dimensionality and it may be vector valued.

An easy way to find a sufficient statistic is to examine the likelihood for the kind of factorization displayed in:

1.5 Lemma

Let the likelihood  $L(\theta \mid D)$  factor so that:

$$L(\theta \mid D) = g(s \mid \theta)h(D) ,$$

where  $h$  does not depend on  $\theta$  . Then  $s$  is a *sufficient statistic* for  $\theta$  .

Proof:

For an arbitrary prior density  $\pi$  , the corresponding posterior density is given by:

$$\begin{aligned} \pi(\theta \mid D) &= L(\theta \mid D)\pi(\theta) / \int_{\mathcal{H}} L(\theta \mid D)\pi(\theta)d\theta \\ &= g(s \mid \theta)h(D)\pi(\theta) / \int_{\mathcal{H}} g(s \mid \theta)h(D)\pi(\theta)d\theta \\ &= g(s \mid \theta)\pi(\theta) / \int_{\mathcal{H}} g(s \mid \theta)\pi(\theta)d\theta . \end{aligned}$$

The last expression clearly depends on  $D$  only through the statistic  $s$  . Thus by Definition 1.4,  $s$  is sufficient for  $\theta$  . ||

For example, if  $x_1, \dots, x_n$  are independent lifelengths in the exponential model, given  $\lambda$  , then  $s \stackrel{\text{def}}{=} \sum_{i=1}^n x_i$  is a sufficient statistic for the failure rate  $\lambda$  since:

$$L(\lambda \mid x_1, \dots, x_n) = \lambda^n \exp \left( -\lambda \sum_{i=1}^n x_i \right) = \lambda^n e^{-\lambda s} .$$

### The Sample Space

The *sample space* is the space or set of possible sample outcomes.

If we observe the lifetimes of  $n$  units, the sample space is

$$(1.7) \quad S = \{(x_1, x_2, \dots, x_n) \mid x_i \geq 0, 1 \leq i \leq n\}.$$

For  $n = 2$ ,  $S$  is the positive quadrant; i.e.,

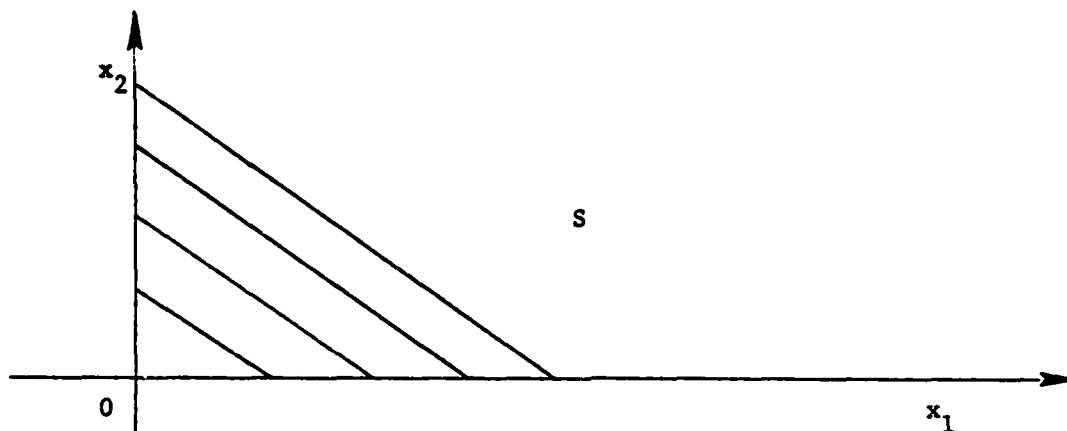


FIGURE 1.2

However, we may just as well consider another sample space. Suppose we are told only the ordered lifetimes

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)},$$

i.e., we no longer know which unit fails at time  $x_{(1)}$ ,  $1 \leq i \leq n$ . The sample space corresponding to the ordered lifetimes is now

$$(1.8) \quad S_0 = \{(x_{(1)}, x_{(2)}, \dots, x_{(n)}) \mid 0 \leq x_{(1)} \leq \dots \leq x_{(n)}\}.$$

For  $n = 2$ , we now have

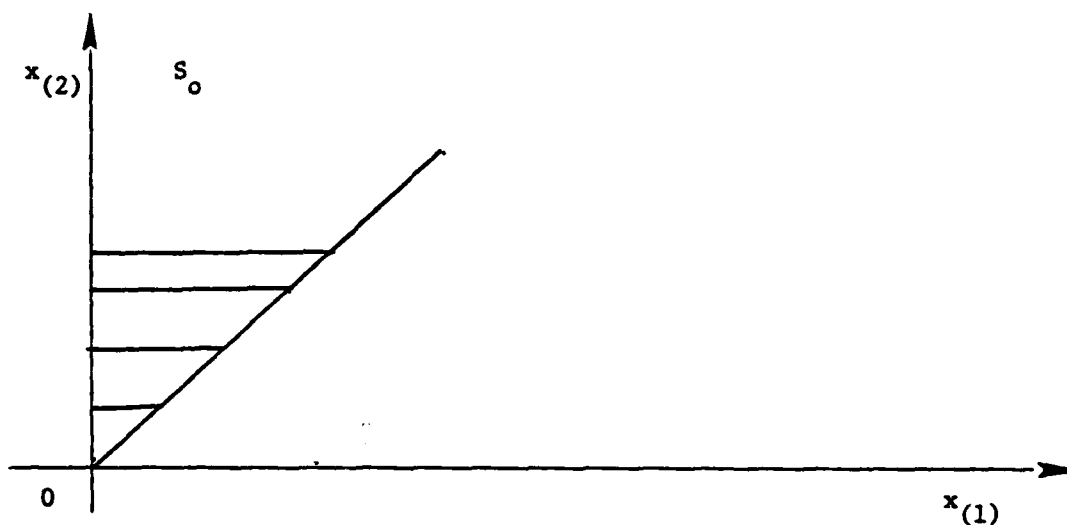


FIGURE 1.3

For sample space (1.7) and the exponential model we have the joint probability density

$$(1.9) \quad p(x_1, x_2, \dots, x_n | \lambda) = \begin{cases} \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} & x_i \geq 0, 1 \leq i \leq n \\ 0 & \text{otherwise} . \end{cases}$$

For this case  $L(\lambda | x_1, x_2, \dots, x_n) = p(x_1, x_2, \dots, x_n | \lambda)$ .

On the other hand, for sample space (1.7) and the exponential model we have

$$(1.10) \quad p_o(x_{(1)}, \dots, x_{(n)} | \lambda) = \begin{cases} n! \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_{(i)} \right\} & 0 \leq x_{(1)} \leq \dots \leq x_{(n)} \\ 0 & \text{otherwise} . \end{cases}$$

The factor  $n!$  in (1.10) follows from the fact that the ordered observations can result from any one of  $n!$  permutations of the observations  $x_1, x_2, \dots, x_n$ . For this case let

$$L_o(\lambda \mid x_{(1)}, \dots, x_{(n)}) = p_o(x_{(1)}, \dots, x_{(n)} \mid \lambda) .$$

From (1.9) and (1.10) we see that

$$(1.11) \quad L(\lambda \mid x_1, x_2, \dots, x_n) = L_o(\lambda \mid x_{(1)}, x_{(2)}, \dots, x_{(n)}) .$$

It follows that, for any prior for  $\tilde{\lambda}$ , the posterior density for  $\tilde{\lambda}$  will be the same no matter which of the above sample spaces we choose. From (1.11) and Lemma 1.5 we see that the order statistics  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ , are sufficient for  $\lambda$ .



## 2. SELECTED LIFE TEST SAMPLING PLANS

In this section we illustrate the application of the concepts and methods of Section 1 when estimating under each of several commonly used life test sampling plans.

### Sampling Plan (a). Complete Observation Until A Specified Number of Failures Have Occurred

A popular plan consists of putting  $n$  items on test and observing the failure times of the first  $k$  failures, where  $k$  is specified in advance. The motivation for following this plan is to save time in determining an estimate of the exponential failure rate.

Let

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(k)}, \quad 1 \leq k \leq n,$$

denote the successive times of failure of the earliest  $k$  failures;

$x_{(1)}, \dots, x_{(k)}$  are called the first  $k$  *order statistics* in a sample of size  $n$ . The likelihood of this observed outcome under the exponential model is given by;

$$L(\lambda \mid D) = \frac{n!}{1! \dots 1!(n-k)!} \left[ \prod_{i=1}^k \lambda e^{-\lambda x_{(i)}} \right] e^{-(n-k)\lambda x_{(k)}}.$$

(We follow the usual convention that  $0! \equiv 1$ .)

Simplifying, we have:

$$(2.1) \quad L(\lambda \mid D) = \frac{n!}{(n-k)!} \lambda^k \exp \left\{ -\lambda \left[ \sum_{i=1}^k x_{(i)} + (n-k)x_{(k)} \right] \right\}.$$

To verify the expression preceding (2.1), note that the combinatorial

coefficient  $\frac{n!}{1! \dots 1!(n-k)!}$  represents the number of ways of choosing

one observation to correspond to each of  $x_{(1)}, x_{(2)}, \dots, x_{(k)}$  and  $(n - k)$  observations for the  $(n - k)$  unobserved failure times, from among the  $n$  failure times (of which only the first  $k$  are actually observed). The product factor represents the joint density of the  $k$  actually observed failure times, given  $\lambda$ . Finally, the last factor represents the probability that  $(n - k)$  lifelengths each exceed  $x_{(k)}$ , given  $\lambda$ .

The expression in the exponent of (2.1):

$$(2.2) \quad \sum_{i=1}^k x_{(i)} + (n - k)x_{(k)} = nx_{(1)} + (n - 1)(x_{(2)} - x_{(1)}) \\ + \dots + (n - k + 1)(x_{(k)} - x_{(k-1)}) \stackrel{\text{def}}{=} T(x_{(k)})$$

represents the *total time on test until the  $k^{\text{th}}$  failure*. Note that it is comprised of  $nx_{(1)}$  the total time on test observed until the first failure; of  $(n - 1)(x_{(2)} - x_{(1)})$ , the total time on test observed between the first and second failures; ... ; and of  $(n - k + 1)(x_{(k)} - x_{(k-1)})$ , the total time on test observed between the penultimate and the last observed failures. (Of course, it is understood that after an item fails, it is no longer under observation.)

The total time on test statistic turns out to be a very important and useful statistic not only in the exponential model, but also after appropriate generalization, in a large number of other models involving incomplete data. In the exponential model, the total time on test and the number of observed failures constitute a sufficient statistic for  $\lambda$ . In the presently considered sampling plan under which  $k$ , the number of observed failures is specified in advance, we see from (2.1) that  $T(x_k)$  alone is sufficient for  $\lambda$ .

Suppose we assume a gamma prior on  $\lambda$ :

$$\pi(\lambda) = b^a \lambda^{a-1} e^{-b\lambda} / \Gamma(a) .$$

Using (2.1), we obtain for the posterior:

$$(2.3) \quad \pi(\lambda \mid D) = [b + T(x_{(k)})]^{k+a} \lambda^{k+a-1} e^{-\lambda[b+T(x_{(k)})]} / \Gamma(k+a) ,$$

also a gamma, but with the shape parameter  $a$  of the prior density replaced by  $a + k$  and the scale parameter  $b$  of the prior density replaced by  $b + T(x_{(k)})$ . Note that the increment in the scale parameter is  $T(x_{(k)})$ , the observed total time on test. The mode of the posterior density is:

$$\lambda^0 = (k + a - 1) / [b + T(x_{(k)})] .$$

It is interesting to note that for  $\pi(\lambda) \equiv c$ , the mode of the posterior is exactly the well known MLE:

$$\hat{\lambda} = k / T(x_{(k)}) .$$

However, it should be emphasized that this prior is *improper* in the sense

$$\text{that } \int_0^{\infty} \pi(\lambda) d\lambda = \infty .$$

We would expect that after collecting a set of data from the exponential distribution, we would have more information concerning the unknown parameter  $\lambda$  than before; more precisely, the peakedness of the density of  $\lambda$  might increase or the coefficient of variation might decrease. The *coefficient of variation* of a distribution is the ratio of the standard deviation (assumed finite) to the absolute value of the mean (assumed non-zero). In the case of the gamma prior density, the mean is  $a/b$ , the variance is  $a/b^2$ , and so the corresponding coefficient of variation is  $a^{-1/2}$ .

Under the present sampling plan, the posterior distribution, given in (2.3), is also a gamma distribution, but with the shape parameter

(the updated "a" value) now  $a + k$ . It follows that the coefficient of variation is now reduced to  $(a + k)^{-1/2}$ . Thus for fixed  $a$ , the coefficient of variation decreases roughly as  $k^{-1/2}$ , where  $k$  denotes the observed number of failures. This simple calculation gives us a quantitative notion as to our relative uncertainty concerning  $\lambda$  both before and after observation, and therefore concerning the decrease in our uncertainty.

Sampling Plan (b). Observation Terminated at Fixed Time (Truncated Sampling)

Suppose  $n$  units are put on life test at time  $t = 0$  and each is observed until failure or fixed time  $t_0$ , whichever occurs first. Given the random outcome  $K = k$  ( $0 \leq k \leq n$ ) observed failures in  $[0, t_0]$ , the corresponding likelihood is given by:

$$(2.4) \quad L(\lambda \mid D) = \frac{n!}{1! \dots 1!(n-k)!} \left[ \prod_{i=1}^k \lambda e^{-\lambda x_{(i)}} \right] e^{-\lambda(n-k)t_0},$$

where the product in square brackets is defined as 1 for  $k = 0$ . The verification of the likelihood expression (2.4) is similar to that obtained under the previous sampling plan, leading to the expression preceding (2.1). One key difference is that the number  $n - k$  surviving fixed time  $t_0$  under the present plan is random, while the number  $n - k$  surviving past the  $k^{\text{th}}$  failure under the earlier plan is specified in advance.

We may rewrite the likelihood  $L(\lambda \mid D)$  as:

$$(2.5) \quad L(\lambda \mid D) = \frac{n!}{(n-k)!} \lambda^k \exp \left\{ -\lambda \left[ \sum_{i=1}^k x_{(i)} + (n-k)t_0 \right] \right\}.$$

It is clear from (2.5) and Lemma 1.5 that the pair  $k$  and the total time on test:

$$T(t_0) = \sum_{i=1}^k x_{(i)} + (n - k)t_0$$

are sufficient statistics for  $\lambda$ . Under the present sampling plan, both  $k$  and the  $T(t_0)$  are observed, and thus together constitute the data  $D$ .

From (2.5) the MLE is computed as:

$$\hat{\lambda} = k/T(t_0).$$

Under the present sampling plan it is possible to observe  $k = 0$  failures, leading to a MLE of 0. Such an estimate is intuitively unsatisfactory; in this situation an analysis based on the posterior distribution of  $\lambda$  is preferable.

The posterior density resulting from a gamma prior [see (1.4)] is, using (2.5):

$$(2.6) \quad \pi(\lambda | D) = [b + T(t_0)]^{k+a} \lambda^{k+a-1} e^{-\lambda[b+T(t_0)]} / \Gamma(k+a).$$

Just as in the previous sampling plan, the posterior coefficient of variation is  $(a+k)^{-1/2}$ , which is smaller than  $a^{-1/2}$ , the prior coefficient of variation.

#### Sampling Plan (c). Inverse Binomial Sampling

A unit is put on test until it fails or reaches a specified age  $t_0$ , whichever occurs first. At this time, the unit is replaced by a new unit. This procedure is repeated sequentially until  $k$  (specified

in advance) failures are actually observed. The number  $N$  of units that have to be tested until  $k$  failures are actually observed is, of course, a random variable. For example, we may have only one test chamber and we may be able to test at most one unit at a time.

Let  $Y = \min(X, t_0)$ , where  $X$  has exponential density  $\lambda e^{-\lambda x}$ ,  $x \geq 0$ . Then conditional on  $Y < t_0$  and  $T$ , the density of  $Y$  is given by:

$$\lambda e^{-\lambda y} / (1 - e^{-\lambda t_0}) \quad \text{for } 0 \leq y \leq t_0.$$

Thus, if the successive failure ages actually observed are denoted by  $y_1, y_2, \dots, y_k$ , the corresponding conditional joint density is given by:

$$g(y_1, \dots, y_k \mid \lambda, y_i < t_0, 1 \leq i \leq k) = \prod_{i=1}^k \frac{\lambda e^{-\lambda y_i}}{(1 - e^{-\lambda t_0})} = \frac{\lambda^k \exp\left(-\lambda \sum_{i=1}^k y_i\right)}{(1 - e^{-\lambda t_0})^k}.$$

The probability that  $N = n$  units have to be tested in order to observe  $k$  actual failures is given by

$$P[N = n \mid \lambda] = \binom{n-1}{k-1} (1 - e^{-\lambda t_0})^k e^{-\lambda t_0(n-k)} \quad \text{for } n \geq k.$$

It follows that, given  $N = n$  and observed failure ages  $y_1, y_2, \dots, y_k$ , the corresponding likelihood is:

$$L(\lambda \mid D) = \binom{n-1}{k-1} \lambda^k \exp\left(-\lambda \sum_{i=1}^k y_i\right) \exp[-\lambda t_0(n-k)], \quad 0 < y_i < t_0, \quad i = 1, \dots, k.$$

Combining exponentials we obtain:

$$(2.7) \quad L(\lambda \mid D) = \binom{n-1}{k-1} \lambda^k \exp \left\{ -\lambda \left[ \sum_{i=1}^k y_i + (n-k)t_0 \right] \right\},$$

$$0 < y_i < t_0, \quad i = 1, \dots, k.$$

From Lemma 1.5, we conclude that *the total time on test*:

$$T = \sum_{i=1}^k y_i + (n-k)t_0$$

is by itself sufficient for  $\lambda$  since  $k$  is fixed in advance.

From (2.7) we also obtain the posterior density for  $\lambda$  based on a gamma prior density as:

$$(2.8) \quad \pi(\lambda \mid D) = (b+T)^{k+a} \lambda^{k+a-1} e^{-\lambda(b+T)} / \Gamma(k+a).$$

Note the striking resemblance in the posterior density under the three sampling plans so far considered! (Compare (2.8), (2.6), and (2.3).)

Remarks:

In the three sampling plans considered so far, the number  $k$  of observed failures and the total time on test  $T$  are all that we need from the observations in order to complete our data analysis; the *sufficiency* of  $k$  and  $T$  makes this true. Note that this fact holds for any choice of a prior density.

The "stopping rule" in each of the three test plans considered gives no information about the parameter  $\lambda$ , i.e., is noninformative about  $\lambda$ . (As the name implies, the *stopping rule* is simply the rule which specifies when the testing is to stop. For example, in the sampling plan (a) we test until  $k$  failures are observed and then stop further observation.) If the stopping rule were to give information

about the parameter  $\lambda$ , then the total time on test  $T$  and the observed number of failures would *not* be sufficient for  $\lambda$ .

Finally, note that in the test plans considered thus far, the MLE is the ratio of the number of observed failures to the total time on test. This simple formula for the MLE holds in most of the testing procedures followed under the exponential model, as we see in this chapter and shall see in Chapter 2.



### 3. INFERENCE BASED ON MEAN LIFE

Thus far we have discussed inference for the exponential distribution based on the failure rate  $\lambda$ . For many analysts, the mean life  $\theta$  of the exponential distribution may seem to be the more appropriate parameter to estimate. Note that either parameter determines the other in the exponential model since:

$$\theta = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \lambda^{-1}.$$

Suppose we consider the simplest type of testing plan:  $n$  units are put on test and observed until each fails. The corresponding mutually independent lifelengths given  $\lambda$  are  $x_1, x_2, \dots, x_n$ , constituting a complete sample from the exponential density:

$$f(x | \theta) = \theta^{-1} e^{-x/\theta}.$$

The *likelihood* of the outcome is given by:

$$(3.1) \quad L(\theta | x_1, \dots, x_n) = \theta^{-n} \exp \left( -\theta^{-1} \sum_{i=1}^n x_i \right).$$

Clearly, the MLE of  $\theta$  is given by:

$$\hat{\theta} = n^{-1} \sum_{i=1}^n x_i.$$

Suppose now we have very little prior information on the parameter  $\theta$ . We therefore assume a *rectangular prior density* on  $\theta$ :

$$\pi(\theta) = M^{-1} \quad \text{for } 0 \leq \theta \leq M,$$

where  $M$  is large.<sup>†</sup> The corresponding *posterior density* for  $\tilde{\theta}$  may be computed approximately, by Bayes' Theorem (Theorem 1.2), as:

$$(3.2) \quad \pi(\theta \mid x_1, \dots, x_n) = b^a \theta^{-(a+1)} e^{-b/\theta} / \Gamma(a)$$

for  $\theta, a, b > 0$ , where now  $a = n - 1$  and  $b = \sum_{i=1}^n x_i$ .

The density of (3.2), denoted by  $\pi_{a,b}(\theta)$ , is called the *inverted gamma density*, since if  $\tilde{\theta}$  is a random variable with density  $\pi_{a,b}(\theta)$ , then  $\tilde{\theta}^{-1} = \tilde{\lambda}$  has gamma density (1.5).

We may verify readily that if our initial prior is of the form  $\pi_{a,b}(\theta)$ , of (3.2) and we use any one of the sampling plans, then the corresponding posterior density is also of the form (3.2). However, the parameter  $a$  of the prior is replaced by  $a + k$  in the posterior, and the parameter  $b$  of the prior is replaced by  $b + T$  in the posterior. As before,  $k$  denotes the number of observed failures and  $T$  is the total time on test. This follows readily from the likelihood expression (3.1) and the fact that in this case  $T = \sum_{i=1}^n x_i$  since all lifelengths are observed.

#### The Mean of the Posterior Density

The mean of the inverted gamma density given in (3.2) is readily

<sup>†</sup> It would be inconsistent mathematically to assume a rectangular prior on both  $\tilde{\lambda}$  and  $\tilde{\theta} = \tilde{\lambda}^{-1}$ , even though we have very little prior information on both parameters. (The implications for  $\tilde{\theta}$  that result from assuming a rectangular prior for  $\tilde{\lambda}$  are brought out in Exercise 1.) We assume a rectangular prior for  $\tilde{\theta}$  here to motivate the use of a natural conjugate prior.

calculated:

$$\int_0^{\infty} \theta \{ b^a \theta^{-(a+1)} e^{-b/\theta} / \Gamma(a) \} d\theta = b/(a-1) .$$

For the inverted gamma posterior density in which the parameters are  $a + k$  (in place of  $a$ ) and  $b + T$  (in place of  $b$ ), the corresponding mean takes the form

$$(3.3) \quad \frac{b + T}{a + k - 1} \equiv (1 - w) \frac{b}{a - 1} + w \frac{T}{k} ,$$

where  $w = k/(k + a - 1)$ . Thus the mean of the posterior density may be written as a convex combination of the prior mean  $b/(a - 1)$  and the maximum likelihood estimate  $T/k$  of the exponential life distribution mean. Note that as  $k$ , the number of observed failures increases, the posterior mean attaches more weight to the MLE of the true mean and less weight to the prior mean.

Table 3.1 summarizes the properties of the natural conjugate prior density and of the corresponding posterior density for the two different parametrizations of the exponential model.

TABLE 3.1  
COMPARISON OF ALTERNATIVE  
PARAMETERIZATIONS OF THE EXPONENTIAL MODEL

Parameter	Failure Rate, $\lambda$	Mean Life, $\theta$
Likelihood	$\lambda^k e^{-\lambda T}$	$\theta^{-k} e^{-T/\theta}$
Natural Conjugate Prior	$b^a \lambda^{a-1} e^{-b\lambda} / \Gamma(a)$ , $a, b > 0$ (Gamma)	$b^a \theta^{-(a+1)} e^{-b/\theta} / \Gamma(a)$ , $a, b > 0$ (Inverted Gamma)
Prior Mean	$a/b$	$b/(a - 1)$ $a > 1$
Prior Variance	$a/b^2$	$b^2 / [(a - 1)^2 (a - 2)]$ $a > 2$
Prior Coefficient of Variation	$a^{-1/2}$	$(a - 2)^{-1/2}$ $a > 2$
Posterior Mean	$(a + k)/(b + T)$	$(b + T)/(a + k - 1)$ $a > 1$
Posterior Variance	$\frac{a + k}{(b + T)^2}$	$\frac{(b + T)^2}{(a + k - 1)^2 (a + k - 2)}$ $a > 2$
Posterior Coefficient of Variation	$(a + k)^{-1/2}$	$(a + k - 2)^{-1/2}$ $a > 2$

Note:  $k$  = number of observed failures,

$T$  = total time on test.

### Arbitrary Prior Density

In our analysis up to now, we have focused mostly on the case in which the prior density is the *natural conjugate prior*. In this subsection, we expand our consideration to cases in which the prior is not necessarily the natural conjugate prior. We obtain results similar to those holding in the natural conjugate prior case.

Let  $\pi(\theta)$  be a prior density on  $(H)$  such that  $\{\theta \mid \pi(\theta) > 0\}$  is an interval on  $[0, \infty)$ . We show in Appendix A of this chapter that

$$(3.4) \quad P[\tilde{\theta} > \theta_0 \mid k, T] = \int_{\theta_0}^{\infty} \pi(\theta \mid k, T) d\theta$$

is *decreasing*<sup>†</sup> in  $k > 0$  for fixed  $T$  and *increasing*<sup>†</sup> in  $T$  for fixed  $k$ , i.e., the posterior random variable  $\tilde{\theta}$  is *stochastically decreasing* in  $k$  and *stochastically increasing* in  $T$ . In particular,

$$E[\tilde{\theta} \mid k = 0, T] \geq E[\tilde{\theta} \mid k = 0, T = 0];$$

the lower bound is, of course, the mean of the prior density. Thus, observing total time on test *without* observing failures tends to change our belief about  $\theta$  as compared with our prior belief; we tend to believe in a larger  $\theta$ . However, for the *natural conjugate prior*, the variance of  $\tilde{\theta}$  given  $k$  and  $T$ , decreases with  $k$  but *increases* with  $T$ . Also the coefficient of variation is *constant* in  $T$  but *decreases with*  $k$ . (See Table 3.1.) Hence if  $k = 0$ , large values of  $T$  tend to make us optimistic regarding  $\theta$ . However, failures are needed to sharpen the posterior density. Similarly, for  $\hat{\lambda} = \tilde{\theta}^{-1}$ , we

<sup>†</sup> *Terminology:* Throughout the book we use the term "increasing" in place of "nondecreasing" and "decreasing" in place of "nonincreasing."

can show that for a general prior density on  $\tilde{\lambda}$  under mild regularity conditions,

$$P[\tilde{\lambda} \geq \lambda_0 \mid k, T]$$

*increases in*  $k$  *for fixed*  $T$  *and decreases in*  $T$  *for fixed*  $k$ , just as we would expect.

### 3.1 Example

Suppose our prior density on  $\tilde{\theta}$  is the natural conjugate prior with  $a = 4$  and  $b = 12$ . We put 10 units on test. The earliest failure occurs at  $X_{(1)} = 1$ , followed by a withdrawal at  $t = 2$ , and finally a second failure at  $X_{(2)} = 3$ . In Figure 3.1, we plot the posterior mean, posterior standard deviation, and posterior coefficient of variation as a function of  $t$ , the test time elapsed. Table 3.1 may be used to generate the plots.

In Figure 3.2 we have plotted the posterior density for  $\tilde{\theta}$  at selected times during the life test. The posterior density for  $t = 0$  is of course the prior density. Notice the shape of the posterior density at  $t = 1^-$  (i.e., just before the first observed failure) and at  $t = 1$  (i.e., just after the first failure).

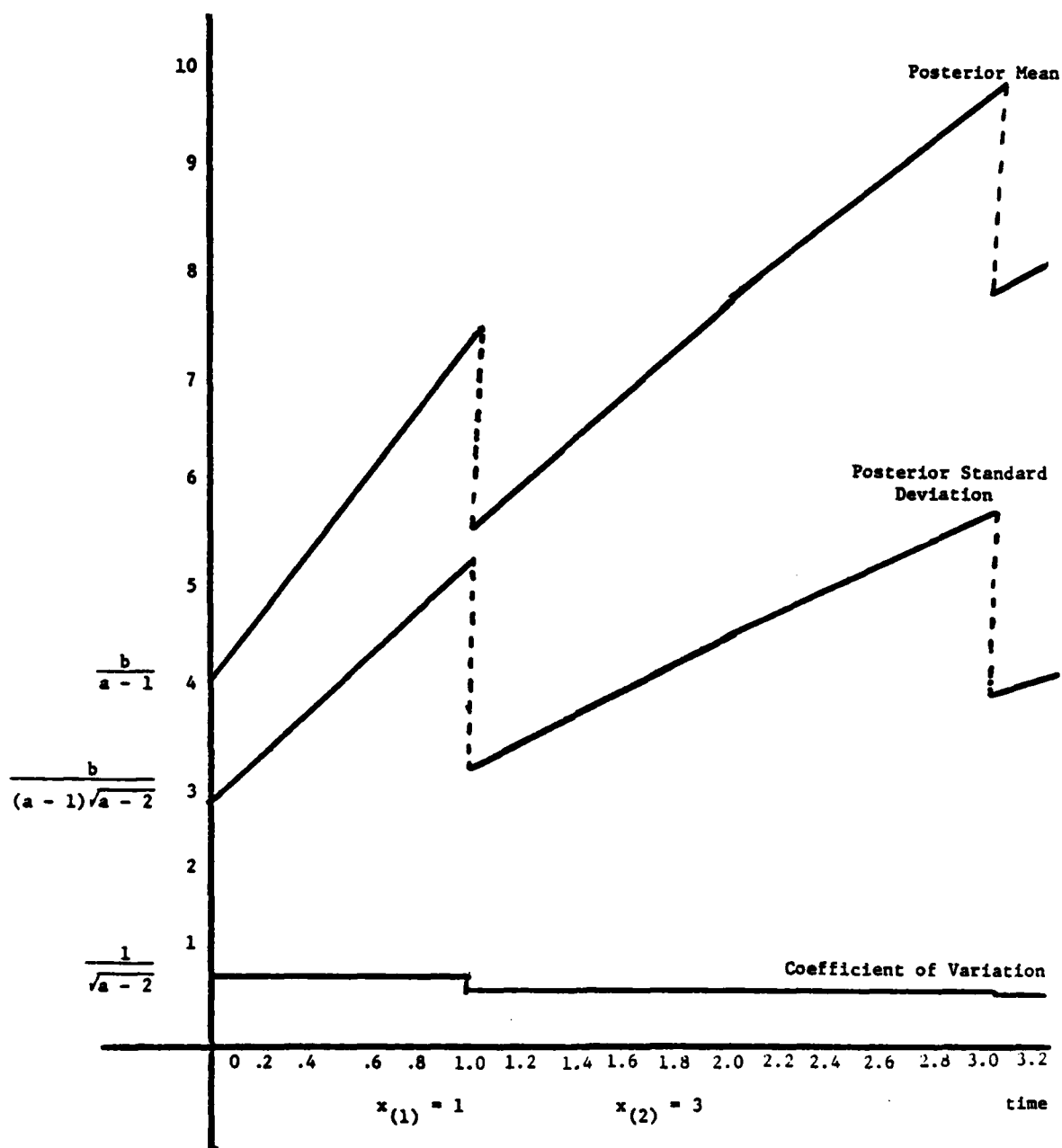


FIGURE 3.1

POSTERIOR MEAN, POSTERIOR STANDARD DEVIATION, AND POSTERIOR COEFFICIENT  
OF VARIATION AS A FUNCTION OF ELAPSED TEST TIME

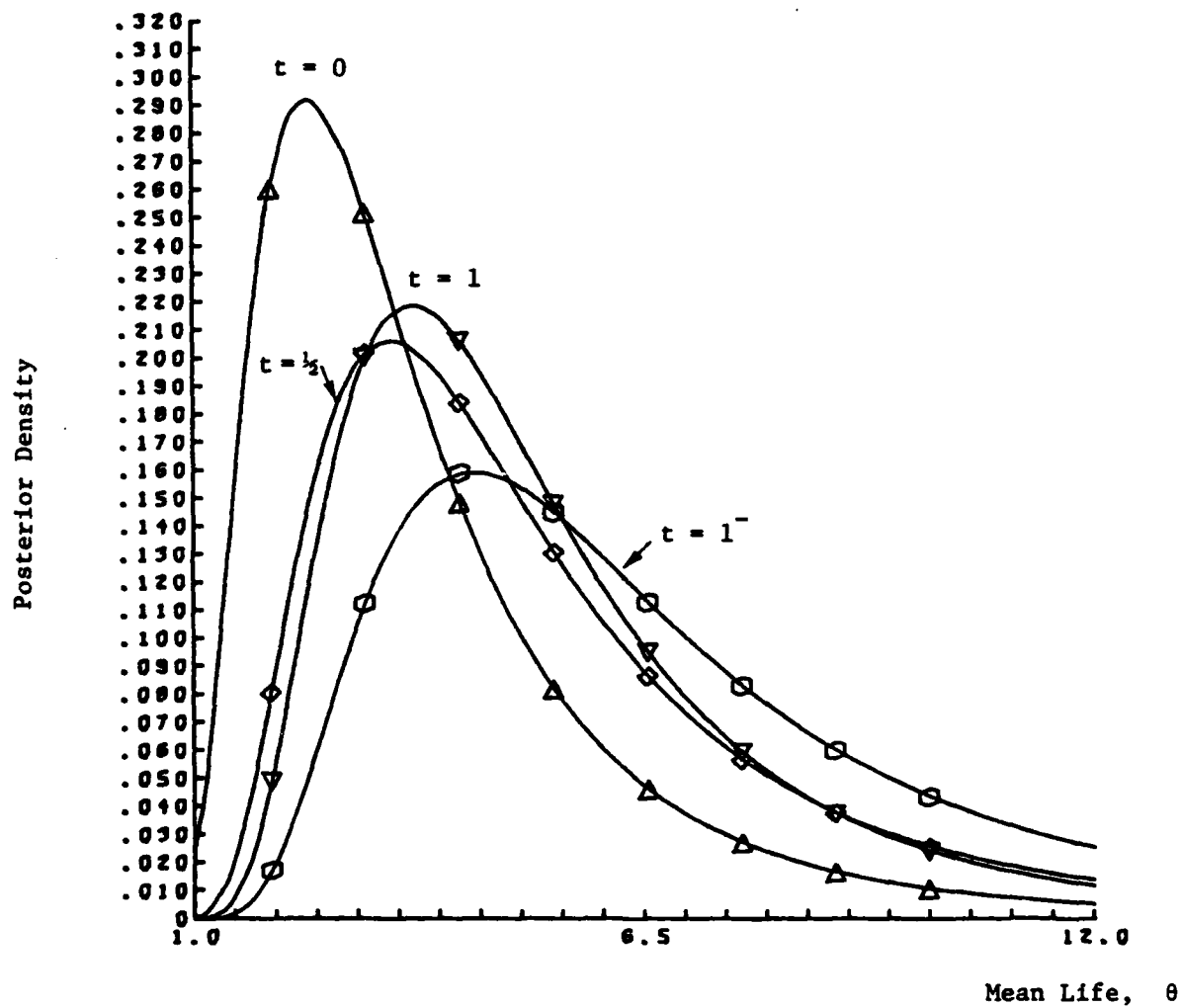


FIGURE 3.2

POSTERIOR DENSITIES AT SELECTED TEST TIMES ( $t$ )



#### 4. NOTES AND REFERENCES

##### Section 1:

Books which emphasize Bayesian concepts as well as their applications are DeGroot (1970) and Winkler (1972). In a series of papers, D. Basu (1975), (1977), (1978), points out the inadequacy of other approaches to statistical inference. Definition 1.4 of sufficiency is attributed by D. Basu to Kolmogorov (1942). The connection between sample theory (Fisher) sufficiency and Kolmogorov sufficiency is discussed in Shih-Chuan Cheng (1978). Fisher sufficiency implies Kolmogorov sufficiency. The converse is false in general. Natural conjugate priors were introduced by Raiffa and Schlaifer (1961). For a rigorous characterization of natural conjugate priors, see Diaconis and Ylvisaker (1979). The reviews of Bayesian statistics by Lindley (1972) and (1978) present excellent summaries of recent advances in the subject.

##### Section 2:

Epstein and Sobel (1953), (1954) were the first to investigate the properties of the exponential model applied to life test plans. In a series of papers they made an intensive and extensive study of the statistical features of a variety of exponential procedures for life testing. Their work greatly influenced reliability theory and practice at the time and still strongly influences current statistical practice in reliability. Government Handbook H-108 and its subsequent modifications represent the Government's "seal of approval" and its effort to implement the theory by making readily available tables and graphs for easy use of the exponential model.

Section 3:

Stochastic monotonicity properties of the posterior mean are derived using the concepts and methods of total positivity. (A comprehensive and authoritative treatment of total positivity may be found in Karlin (1968).) Theorem A.1 in the appendix is similar to Lemma 2, p. 74, of Lehmann (1959).

In the mathematical insurance literature formula (3.3) is called the credibility formula. Jewell (1977) and (1978) has discussed Bayesian life testing.

## REFERENCES

- [1] Barlow, R. E., R. H. Toland and T. Freeman, "Stress-Rupture Life of Kevlar/Epoxy Spherical Pressure Vessels," UCID-17755, Part 3, Lawrence Livermore Laboratories, Livermore, California, 94550, (1979).
- [2] Basu, D., "Statistical Information and Likelihood," Sankhyā, Vol. 37, Series A., Part 1, pp. 1-71, (1975).
- [3] Basu, D., "On the Elimination of Nuisance Parameters," Journal of the American Statistical Association, Vol. 72, pp. 355-366, (1977).
- [4] Basu, D., "On Partial Sufficiency: A Review," Journal of Statistical Planning and Inference, Vol. 2, pp. 1-13, (1978).
- [5] Cheng, Shih-Chuan, "A Mathematical Study of Sufficiency and Adequacy in Statistical Theory," Ph.D. Thesis, Florida State University, Tallahassee, Florida, (1978).
- [6] DeGroot, M. H., OPTIMAL STATISTICAL DECISIONS, McGraw-Hill, New York, (1970).
- [7] Diaconis, R. and D. Ylvisaker, "Conjugate Priors for Exponential Families," Annals of Statistics, Vol. 7, pp. 269-281, (1979).
- [8] Epstein, B. and M. Sobel, "Life Testing," Journal of the American Statistical Association, Vol. 48, pp. 486-502, (1953).
- [9] Epstein, B. and M. Sobel, "Some Theorems Relevant to Life Testing From an Exponential Distribution," Annals of Mathematical Statistics, Vol. 25, pp. 373-381, (1954).
- [10] Handbook H-108, "Sampling Procedures and Tables for Life and Reliability Testing," DoD Document.
- [11] Karlin, S., TOTAL POSITIVITY, Vol. I., Stanford University Press, Stanford, California, (1968).
- [12] Kolmogorov, A. N., "Determination of the Centre of Dispersion and Degree of Accuracy for a Limited Number of Observations," Izvestia Akad. Neuk. U. S. S. R. Ser-Math., Vol. 6, pp. 3-32, (1942), (in Russian).
- [13] Lehmann, E., TESTING STATISTICAL HYPOTHESES, J. Wiley and Sons, New York, (1959).
- [14] Lindley, D. V., "The Bayesian Approach," Scandinavian Journal of Statistics, Vol. 5, pp. 1-26, (1978).
- [15] Lindley, D. V., BAYESIAN STATISTICS, A REVIEW, SIAM, Philadelphia, Pennsylvania, (1972).

- [16] Marshall, A. W. and F. Proschan, "Inconsistency of the Maximum Likelihood Estimator of the Increasing Failure Rate Average Distribution," (in progress).
- [17] Raiffa, H. and R. Schlaifer, APPLIED DECISION THEORY, Graduate School of Business Administration, Harvard University, Boston, Massachusetts, (1961).
- [18] Winkler, R. L., AN INTRODUCTION TO BAYESIAN INFERENCE AND DECISION, Holt, Rinehart and Winston, Inc., New York, (1972).
- [19] Jewell, W. S., "Bayesian Life Testing Using the Total Q on Test," THE THEORY AND APPLICATIONS OF RELIABILITY, Vol. I, edited by C. P. Tsokos and I. N. Shimi, pp. 49-66, Academic Press, (1977).
- [20] Jewell, W. S., "Bayesians Learn While Waiting," The ASTIN Bulletin, Vol. X, Part 2, pp. 163-172, (1978).

## APPENDIX A

POSTERIOR DISTRIBUTIONS CORRESPONDING TO ARBITRARY  
PRIORS AND LIKELIHOODS WITH THE MONOTONE RATIO PROPERTY

We will prove the results mentioned in Section 3.

A.1 Theorem

Let  $\pi_0(\theta)$  be a prior density on  $(H)$  such that  $\{\theta \mid \pi_0(\theta) > 0\}$  is an interval in  $(-\infty, \infty)$ . Let data  $D = (k, T)$  and likelihood  $L(\theta \mid k, T)$  be given such that for  $\theta_1 < \theta_2$

$$L(\theta_1 \mid k, T) / L(\theta_2 \mid k, T)$$

is increasing in  $k$  and decreasing in  $T$ . Let  $g(\theta)$  be increasing in  $\theta$ . Then

$$(A.1) \quad \int_{(H)} g(\theta) \pi(\theta \mid k, T) d\theta$$

is decreasing in  $k$  ( $T$  fixed) and increasing in  $T$  ( $k$  fixed).

The proof will be given shortly.

A.2 Corollary

If  $L(\theta \mid k, T) = \theta^{-k} e^{-T/\theta}$ , then

$$(A.2) \quad P[\bar{\theta} > \theta_0 \mid k, T] = \int_{\theta_0}^{\infty} \pi(\theta \mid k, T) d\theta$$

is decreasing in  $k$  ( $T$  fixed) and increasing in  $T$  ( $k$  fixed). [See Section 3 and (3.4).]

Proof:

In Theorem A.1, let  $g(\theta) \equiv 1$  for  $\theta \geq \theta_0$  and 0 otherwise. ||

A.3 Corollary

Let  $L(\lambda | k, t) \propto \lambda^k e^{-\lambda T}$  and  $g(\lambda)$  be increasing in  $\lambda$ . Then

$$(A.3) \quad \int_0^\infty g(\lambda) \pi(\lambda | k, T) d\lambda$$

is increasing in  $k$  ( $T$  fixed) and decreasing in  $T$  ( $k$  fixed).

Clearly (A.1) [A.3] implies that all moments of the posterior distribution are decreasing [increasing] in  $k$  and increasing [decreasing] in  $T$ .

Proof of Theorem A.1:

For  $\theta_1 < \theta_2$ ,

$$L(\theta_1 | k, T) / L(\theta_2 | k, T)$$

increasing in  $k$  implies that for  $k_1 < k_2$

$$L(\theta | k_1, T) / L(\theta | k_2, T)$$

is increasing in  $\theta$ . It is easy to see that this in turn implies that:

$$(A.4) \quad \pi(\theta | k_1, T) / \pi(\theta | k_2, T)$$

is also increasing in  $\theta$ . Let

$$A = \{\theta | \pi(\theta | k_1, T) > \pi(\theta | k_2, T)\}$$

and

$$B = \{\theta \mid \pi(\theta \mid k_1, T) < \pi(\theta \mid k_2, T)\} .$$

Let  $a = \inf_{\theta \in A} g(\theta)$  and  $b = \sup_{\theta \in B} g(\theta)$  . Then the ratio in (A.4) is

increasing in  $\theta$  implies that  $a \geq b$  .

Now

$$\begin{aligned} & \int_{\Theta} g(\theta) [\pi(\theta \mid k_1, T) - \pi(\theta \mid k_2, T)] d\theta \\ & \geq a \int_A [\pi(\theta \mid k_1, T) - \pi(\theta \mid k_2, T)] d\theta \\ & \quad + b \int_B [\pi(\theta \mid k_1, T) - \pi(\theta \mid k_2, T)] d\theta \\ & = (a - b) \int_A [\pi(\theta \mid k_1, T) - \pi(\theta \mid k_2, T)] d\theta \geq 0 . \end{aligned}$$

Hence, the integral in (A.1) is decreasing in  $k$  .

A similar argument proves that the integral in (A.1) is increasing in  $T$  . ||